METHOD FOR ALLOCATING COMMUNICATION NETWORK RESOURCES USING ADAPTIVE DEMAND PREDICTION

TECHNICAL FIELD

The present invention relates, generally, to Demand Assigned Multiple Access (DAMA) communication networks and, more particularly, to a method for allocating bandwidth within the network using a prediction of future demand based on parameters associated with current users.

BACKGROUND ART AND TECHNICAL PROBLEMS

Demand Assigned Multiple Access (DAMA) communication networks are particularly useful in network applications where traffic is bursty and where broadband service is required, but is not required constantly. DAMA networks allow for the dynamic allocation and reallocation of bandwidth and network resources based on the communication needs of the network users.

Presently known systems allocate bandwidth based on, inter alia, requests for higher priority service from users. For example, if twenty percent of a network's customers are high-priority users, the network may reserve twenty percent of its bandwidth for high-priority use. If, at any given moment, less than twenty percent of the high-priority bandwidth capacity is being utilized, and a low-priority user requests high-priority service, a typical system would grant that request and allow the low-priority user to consume high-priority network resources, since they are available. This can be problematic, particularly in environments with bursty traffic patterns, because the high-priority network resources employed by tow-priority customers may not be available when needed a short time later by a high-priority user.

A method for allocating network resources is thus needed which is capable of predicting future demand to thereby more efficiently allocate network resources among the competing requests from high-priority and low-priority customers.

BRIEF DESCRIPTION OF THE DRAWING

The subject invention will hereinafter be described in conjunction with the appended drawing figure, wherein the reference numerals in the drawing figure correspond to he associated descriptions provided below, and the drawing figure is a schematic block diagram of a method for efficiently allocating network resources based on current user parameters in accordance with a preferred embodiment of the present invention.

DETAILED DESCRIPTION OF THE DRAWING

The efficiency of a Demand Assigned Multiple Access (DAMA) communications network depends on the particular strategies employed in assigning communication resources. Simpler strategies often result in tying up unused resources in order to guarantee quality of service for priority users, which results in inefficiencies associated with the "reserved" resources. To reduce these inefficiencies, the present invention predicts the likelihood of future access demand from users and makes resource allocation decisions based on a combination of the predicted demand and other metrics.

In accordance with a preferred embodiment of the present invention, demand prediction employs the following two

steps: (1) parameters of the statistic distribution of the traffic pattern of each user are adaptably estimated based on the users "duty cycle," i.e., the active and inactive states of the user equipment; and (2) the probability of the active and inactive duration of each user is calculated using the estimated parameters. Other metrics may also be employed in the resource allocation decision, such as user service priority, available network resources and the like.

The drawing figure is a schematic block diagram of the network system 100 including a network resource manager 110 and a plurality of network users represented by respective devices 102, 104, and 106, each of which may include a buffer 107. By way of example, device 102 may be a cellular telephone, device 104 may be a network personal computer (PC), and device 106 may be, for example, a personal digital assistant (PDA) device, all of which compete for bandwidth from the network.

Network resource manager 110 includes a demand history database 112, a demand prediction processor 114, a cost function database 116, a resource pool 118, a decision history database 120 and a network resource allocator 122.

For purposes of this discussion, the time period in which a device is active is denoted as T_{ON} which is followed by a time period in which a device is inactive, denoted as T_{OFF}. The present inventors have determined that, in the aggregate, T_{ON} and as T_{OFF} for traffic in typical data networks often exhibit characteristics of heavily tailed distribution, which may be modeled as for example, a Pareto distribution. Although the techniques described herein work well in the context of a Pareto distribution, the present invention may be employed in the context of virtually any statistical distribution, for example, Weibull, Possion and the like. The Pareto distribution is given by:

Probability density function:	$f(x\|\forall,\exists) = \exists \forall^{\exists} (x)^{-(\exists+1)}$
Cumulative distribution function:	$F(x) = P(X < x) = 1 - (\forall/x)^{\exists}$

Where $\forall$ is the location parameter and $\exists$ is the shape parameter of the distribution.

In accordance with one aspect of the present invention, each device 102, 104 and 106 keeps track of the duration of the T_{ON} and T_{OFF}. The parameters $\forall$ and $\exists$ of the distribution of T_{ON} may be estimated by an adaptive process using each duration sample of T_{ON}. Similarly, the parameters $\forall$ and $\exists$ of the distribution of T_{OFF} may be estimated based on T_{OFF} samples. In the context of the drawing figure, this task may be performed by distribution parameter estimator 108 associated with each device. When a device, for example, device 102, requests network access, it sends the distribution parameters (e.g., $\forall$ and $\exists$) along with the request for network access to network resource manager 110.

Network resource manager 110 stores the distribution parameters of each user, for example, in demand prediction processor 114. When network resource manager 110 receives access requests from active users, it uses the parameters received to estimate the probability $P(T_{ON} > t+)t)$ that the traffic's active period will last longer than a time period t+)t. The probabilities $P(T_{OFF} < t+)t)$ for all inactive users are also updated using the most current saved parameters. The probability calculations may be carried out for multiple values of)t, to thereby allow the predictions of future user traffic to range from, for example, on the order of milliseconds, to seconds or even minutes. Demand prediction processor 114 outputs these estimated probabilities and applies this output to network resource allocator 122.